

# Concentration via chaining method and its applications

Witold Bednorz <sup>\*†‡§</sup>

May 9, 2014

## Abstract

In this paper we study the regularity of paths in terms of properties of admissible nets. We show the right concentration inequality above the modulus of continuity. Using the approach we prove the Bernstein type inequality for the empirical processes. Therefore we obtain the best form of concentration for processes studied recently by Mendelson and Paouris and Tomczak-Jaegerman. Results of this type are of importance in the compressed sensing theory.

## 1 Introduction

In this paper we show how the chaining approach can be used to establish results on the concentration of functionals of random variables. In particular we focus on the well known problem in the theory of empirical processes. We show how to prove the right concentration inequality on the supremum of centered sums of squares of independent random variables. The chaining approach has been recently used to investigate empirical processes [2] and [3]. The main difficulty that appears in the results is to use a special chaining depending on the approximation level. We start our study from the simplest case of a stochastic process with increments under control of one distance. Then we turn to more involved case of two distances. With the results we turn to prove the application to the empirical processes proving the Bernstein type inequality. Finally we discuss how the result can be used in the compressed sensing.

## 2 One distance control of increments

Let  $(T, d)$  be a compact metric space. Let  $X(t)$ ,  $t \in T$  be a stochastic process defined on  $(T, d)$ . We aim to study path properties of  $X(t)$ ,  $t \in T$  under some increment conditions. The simplest setting in which the problem can be analyzed is when there is a single distance  $d$  on  $T$  and a single Young function

---

<sup>\*</sup>**Subject classification:** 60G15, 60G17

<sup>†</sup>**Keywords and phrases:** sample boundedness, Gaussian processes

<sup>‡</sup>Support: Polish Ministry of Science and Higher Education Iuventus Plus Grant no. IP 2011 000171

<sup>§</sup>Institute of Mathematics, University of Warsaw, Banacha 2, 02-097 Warszawa, Poland

$\psi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ , i.e. convex, increasing and such that  $\psi(0) = 0$  and  $\psi(1) = 1$  which are used to impose the following control on the increments

$$\mathbf{E}\psi\left(\frac{|X(t) - X(s)|}{d(s,t)}\right) \leq 1, \text{ for all } s, t \in T. \quad (1)$$

Note that defining for all  $s, t \in T$

$$\|X(t) - X(s)\|_\psi = \inf\{C > 0 : \mathbf{E}\psi\left(\frac{|X(t) - X(s)|}{C}\right) \leq 1\} < \infty \quad (2)$$

one can use distance  $d$  of the form  $d(s, t) = \|X(t) - X(s)\|_\psi$ . Obviously the condition (1) implies some concentration inequality for increments, i.e.

$$\mathbf{P}(|X(t) - X(s)| \geq ud(s, t)) \leq \mathbf{E}\psi\left(\frac{|X(t) - X(s)|}{d(s, t)}\right)/\psi(u) \leq \frac{1}{\psi(u)} \text{ for } u > 0.$$

Note that the requirement that  $\psi$  is convex can be slightly relaxed to the condition that  $\psi$  is continuous, increasing to infinity,  $\psi(0) = 0$  and  $\psi(1) = 1$ . It usually concerns the case when the convexity starts for large enough arguments.

The simplest example of such an increment control is when  $\psi$  is exponential, i.e. we assume that  $\psi(x) = \varphi_p(x) = 2^{x^p} - 1$ ,  $p > 0$ . Then (1) is equivalent to the following concentration inequality

$$\mathbf{P}(|X(t) - X(s)| > d(s, t)u) \leq A \exp(-Bu^p), \text{ for all } u \geq 0,$$

where  $A, B$  - universal constants. In order to analyze the whole class of exponential functions  $\psi$  we formulate the general exponential-type condition such functions should satisfy

$$\psi^{-1}(xy) \leq K(\psi^{-1}(x) + \psi^{-1}(y)) \quad x, y \geq 0, \quad (3)$$

where  $K < \infty$  is a universal constant. Note that  $K = 1$  for  $\psi = \varphi_p$ ,  $p \geq 1$ , nevertheless the inequality holds for all  $\psi = \varphi_p$ ,  $p > 0$  but for  $p < 1$ ,  $K$  depends on  $p$ .

In the particular case of centered Gaussian processes  $X(t)$ ,  $t \in T$  one can apply  $\psi(x) = \varphi_2(x) = 2^{x^2} - 1$  and  $d(s, t) = \sqrt{8/(3 \ln 2)}(\mathbf{E}|X(t) - X(s)|^2)^{\frac{1}{2}}$ . The meaning of (1) is that we acknowledge Gaussian-type concentration for increments, as the best possible tool to understand path properties of the process.

### 3 The concentration inequality for one distance

We start our study towards concentration inequalities by the chaining approach from the simplest setting of a compact  $(T, d)$  and a single Young function  $\psi$ . It should be stressed that if  $T$  is not a completely bounded space in  $d$  one can construct processes that satisfies (1) and which are not sample bounded.

In this setting there exists a separable modification of  $X(t)$ ,  $t \in T$  which we refer to from now on. Indeed (1) implies the continuity in probability of the process and since  $(T, d)$  is compact we can define a separable modification of  $X(t)$ ,  $t \in T$  based on any separable dense subset  $T_0 \subset T$ .

Let  $(T, d)$  be a compact metric space. Let  $N_n = \psi(2^n)$ , for  $n \geq 1$  and  $N_0 = 1$ . We say that a sequence of finite sets  $(T_n)_{n=0}^\infty$  is admissible, if  $|T_n| \leq N_n$ ,  $T_n \subset T_{n+1}$  and  $\bigcup_{n=0}^\infty T_n$  is dense in  $T$ .

For each  $n \geq 0$  define  $\pi_n(t) \in T_n$  in a way that  $d(t, T_n) = d(t, \pi_n(t))$ . The first question we study is what one can say on the difference  $|X(t) - X(\pi_m(t))|$  for a given  $m \geq 0$ . For each  $m \geq 0$  define

$$\sigma_m(t) = \sum_{n=m}^{\infty} 2^n d(t, T_n).$$

Moreover let

$$Z = \sum_{n=1}^{\infty} \frac{1}{N_n^3} \sum_{t \in T_n} \psi\left(\frac{|X(t) - X(\pi_{n-1}(t))|}{d(t, \pi_{n-1}(t))}\right).$$

Clearly  $\mathbf{E}Z \leq 1$ , since  $\frac{N_n^2}{N_n^3} \leq \frac{1}{N_n} \leq \frac{1}{2^n}$ , by the convexity of  $\psi$ . The following result is a stronger version of the usual concentration inequality obtained by the chaining approach (cf. [4]).

**Proposition 1** *For each separable  $X(t)$ ,  $t \in T$  that satisfies (1) for  $\psi$  that verifies the exponential condition (3) the following holds. For all  $t \in T$  and  $m \geq 0$*

$$|X(t) - X(\pi_m(t))| \leq A\sigma_m(t) + Bd(t, \pi_m(t))\psi^{-1}(Z),$$

where  $A = 15K^2$ ,  $B = 4K^2$  are universal constants,  $Z \geq 0$  and  $\mathbf{E}Z \leq 1$ .

**Proof.** Fix  $t \in T$ . We define  $n_i = n_i(t)$ ,  $i \geq 0$  that may depend on  $t$  in the following way. Let  $n_0 = m$  and for  $i \geq 1$  let  $n_i = \inf\{n > n_{i-1} : 2d(t, T_n) < d(t, T_{n_{i-1}})\}$ . We use the following chaining argument

$$|X(t) - X(\pi_m(t))| \leq \sum_{i=1}^{\infty} |X(\pi_{n_i}(t)) - X(\pi_{n_{i-1}}(t))|, \quad (4)$$

Consequently due to (3)

$$\begin{aligned} & |X(\pi_{n_i}(t)) - X(\pi_{n_{i-1}}(t))| \leq \\ & \leq d(\pi_{n_i}(t), \pi_{n_{i-1}}(t))\psi^{-1}\left(\frac{N_{n_i}^3}{N_{n_i}^3}V_{n_i}\right) = \\ & \leq d(\pi_{n_i}(t), \pi_{n_{i-1}}(t))[3K^22^{n_i} + K^2\psi^{-1}\left(\frac{1}{N_{n_i}^3}V_{n_i}\right)], \end{aligned}$$

where

$$V_n = \sum_{u, v \in T_n} \psi\left(\frac{|X(u) - X(v)|}{d(u, v)}\right) \text{ for all } n \geq 0.$$

By the definition of  $n_i$ ,  $i \geq 0$

$$2d(t, T_{n_{i-1}}) \geq d(t, T_{n_{i-1}}).$$

Therefore

$$\begin{aligned} d(\pi_{n_i}(t), \pi_{n_{i-1}}(t)) & \leq d(t, \pi_{n_i}(t)) + \\ & + d(t, \pi_{n_{i-1}}(t)) \leq d(t, T_{n_i}) + 2d(t, T_{n_{i-1}}) \end{aligned}$$

and hence

$$\begin{aligned} 3K^2 \sum_{i=1}^{\infty} 2^{n_i} d(\pi_{n_i}(t), \pi_{n_{i-1}}(t)) &\leq 3K^2 \sum_{i=1}^{\infty} 2^{n_i} (d(t, T_{n_i}) + 2d(t, T_{n_{i-1}})) \leq \\ &\leq 3K^2 \left( \sum_{n=0}^{\infty} 2^n d(t, T_n) + 4 \sum_{n=0}^{\infty} 2^n d(t, T_n) \right) = 15K^2 \sigma_m(t). \end{aligned}$$

On the other hand

$$\begin{aligned} \sum_{i=1}^{\infty} d(\pi_{n_i}(t), \pi_{n_{i-1}}(t)) &\leq 2 \sum_{i=0}^{\infty} d(t, T_{n_i}) \leq \\ &\leq 2 \left( \sum_{i=0}^{\infty} 2^{-i} \right) d(t, T_m) \leq 4d(t, T_m) \end{aligned}$$

and  $\sum_{i=0}^{\infty} \frac{1}{N_{n_i}^3} V_{n_i} \leq Z$ . Thus we finally get

$$|X(t) - X(\pi_m(t))| \leq 15K^2 \sigma_m(t) + 4K^2 d(t, T_m) \psi^{-1}(Z).$$

It completes the proof with  $A = 15K^2$  and  $B = 4K^2$ . ■

**Remark 1** Note that the above proof works for any  $\psi$  that is not necessarily convex, however then the constants  $A, B$  may depend on  $\psi$  since there may be no longer true that  $\psi(2^i) \geq 2^i$ . In particular for  $\psi(x) = \varphi_p(x) = 2^{x^p} - 1$ ,  $0 < p \leq 1$  we have  $K = 2^{\frac{1}{p}-1}$  and  $\mathbf{E}Z \leq (2^{\frac{1}{p}} - 1)^{-1}$ . Consequently  $Z$  has to be multiplied by  $2^{\frac{1}{p}} - 1$  which affects  $A, B$ .

**Corollary 1** Under assumptions of Proposition 1, there exist universal constants  $A, B$  such that

$$\mathbf{E} \sup_{m \geq 0} \sup_{t \in T} \psi \left( \frac{(|X(t) - X(\pi_m(t))| - A\sigma_m(t))_+}{Bd(t, \pi_m(t))} \right) \leq 1.$$

Now we turn to analyze the modulus of continuity of  $X(t)$ ,  $t \in T$ . Consider any  $s, t \in T$ . There exists the smallest  $k \geq 0$  such that both  $d(s, T_k)$  and  $d(t, T_k)$  are less than  $d(s, t)$ , i.e.

$$k(s, t) = \max\{k \geq 0 : d(s, T_k) + d(t, T_k) \geq d(s, t)\}.$$

Note that  $k(s, t)$  is well defined since at least  $d(s, T_0) + d(t, T_0) \geq d(s, t)$ . Obviously  $k(s, t) + 1$  is the required level where in the chaining construction it is better to jump from the approximation of  $t$  to the approximation of  $s$ . We explain proving that  $k(s, t)$  the value  $k \geq 0$  for which the function

$$f(k) = \sigma_{k+1}(s) + \sigma_{k+1}(t) + (2^{k+1} - 1)d(s, t)$$

is the smallest possible. Indeed  $f(k) > f(k-1)$  implies that

$$2^k(d(s, T_k) + d(t, T_k)) > 2^k d(s, t)$$

and hence  $k > k(s, t)$ . On the other hand for all  $k \leq k(s, t)$ , we have that  $f(k) \leq f(k-1)$ . Therefore  $k(s, t)$  is the argument minimum of  $f$ . Consequently in the view of the proof of Proposition 1 the deterministic part that bounds  $|X(t) - X(s)|$  should be up to a constant bounded by

$$\bar{\tau}(s, t) = f(k(s, t)) = \sigma_{k(s, t)+1}(s) + \sigma_{k(s, t)+1}(t) + 2^{k(s, t)+1}d(s, t).$$

We are ready to define a simple distance  $\tau$  on  $T$  such that  $\tau(s, t)$  is comparable with this quantity.

Define the following numbers

$$\sigma(t, a) = \sum_{n=0}^{\infty} 2^n \min\{d(t, T_n), a\}, \quad a > 0, \quad t \in T.$$

Let  $\tau$  be a new distance on  $T$  given by

$$\tau(s, t) = \max\{\sigma(t, d(s, t)), \sigma(s, d(s, t))\}, \quad \text{for all } s, t \in T.$$

We show that  $\tau(s, t)$  is indeed a distance on  $T$  i.e. it satisfies the triangle inequality. It suffices to check that for  $s, t, u \in T$  there holds

$$\tau(s, u) \leq \tau(s, t) + \tau(t, u).$$

First note that

$$\tau(s, u) \leq \max\{\sigma(s, d(s, t) + d(t, u)), \sigma(u, d(s, t) + d(t, u))\}.$$

and then

$$\begin{aligned} \sigma(s, d(s, t) + d(t, u)) &\leq \sum_{n=0}^{\infty} 2^n d(s, T_n) 1_{d(s, T_n) \leq d(s, t)} + \\ &+ \sum_{n=0}^{\infty} 2^n \min\{d(s, t) + d(t, T_n), d(s, t) + d(t, u)\} 1_{d(s, T_n) > d(s, t)} \leq \\ &\leq \sum_{n=0}^{\infty} 2^n \min\{d(s, t), d(s, T_n)\} + \sum_{n=0}^{\infty} 2^n \min\{d(t, u), d(t, T_n)\} \leq \tau(s, t) + \tau(t, u). \end{aligned}$$

In the same way we get

$$\sigma(u, d(s, t) + d(t, u)) \leq \tau(s, t) + \tau(t, u).$$

The distance  $\tau(s, t)$  is comparable with  $\bar{\tau}(s, t)$ , namely

$$2^{-1}\bar{\tau}(s, t) \leq \tau(s, t) \leq \bar{\tau}(s, t). \quad (5)$$

Indeed since for  $k = k(s, t)$  we have  $d(s, T_l) + d(t, T_l) \geq d(s, t)$  for  $l \leq k$  it implies that

$$\begin{aligned} \tau(s, t) &\geq \max\{\sigma_{k+1}(s) + \sum_{l=0}^k 2^l (d(s, T_l) \wedge d(s, t)), \sigma_{k+1}(t) + \sum_{l=0}^k 2^l (d(t, T_l) \wedge d(s, t))\} \geq \\ &\geq 2^{-1}[\sigma_{k+1}(s) + \sigma_{k+1}(t) + 2(2^{k+1} - 1)d(s, t)] = 2^{-1}\bar{\tau}(s, t). \end{aligned}$$

On the other hand

$$\begin{aligned}\tau(s, t) &\leq \max\{\sigma_{k+1}(s), \sigma_{k+1}(t)\} + \sum_{l=0}^k 2^l d(s, t) \leq \\ &\leq \sigma_{k+1}(s) + \sigma_{k+1}(t) + 2^{k+1} d(s, t) = \bar{\tau}(s, t).\end{aligned}$$

We recall that

$$Z = \sum_{n=1}^{\infty} \frac{1}{N_n^3} \sum_{t \in T_n} \psi\left(\frac{|X(t) - X(\pi_{n-1}(t))|}{d(t, \pi_{n-1}(t))}\right).$$

Let us state the main result of this section that  $\tau(s, t)$  is the suitable modulus of continuity.

**Proposition 2** *For each admissible net  $(T_n)_{n=0}^{\infty}$ , and separable  $X(t)$ ,  $t \in T$  that satisfies (1) the following inequality holds. For all  $s, t \in T$*

$$|X(s) - X(t)| \leq \bar{A}\tau(s, t) + \bar{B}d(s, t)\psi^{-1}(Z),$$

where  $\bar{A} = 30K^2$ ,  $\bar{B} = 10K^2$ ,  $Z \geq 0$  and  $\mathbf{E}Z \leq 1$ .

**Proof.** Let  $k = k(s, t)$  for  $s, t \in T$ . Observe that

$$\begin{aligned}|X(t) - X(s)| &\leq |X(t) - X(\pi_{k+1}(t))| + \\ &+ |X(s) - X(\pi_{k+1}(t))| + |X(\pi_{k+1}(t)) - X(\pi_{k+1}(s))|.\end{aligned}\tag{6}$$

By Proposition 1 we have that for  $x \in \{s, t\}$

$$|X(x) - X(\pi_{k+1}(x))| \leq A\sigma_{k+1}(x) + Bd(x, \pi_{k+1}(x))\psi^{-1}(Z),$$

where  $A = 15K^2$  and  $B = 4K^2$ . Since

$$d(s, \pi_{k+1}(s)) + d(t, \pi_{k+1}(t)) \leq d(s, t)$$

it implies that

$$\begin{aligned}|X(s) - X(t) - X(\pi_{k+1}(s)) + X(\pi_{k+1}(t))| &\leq \\ &\leq A(\sigma_{k+1}(s) + \sigma_{k+1}(t)) + Bd(s, t)\psi^{-1}(Z).\end{aligned}\tag{7}$$

On the other hand using our main argument

$$\begin{aligned}|X(\pi_{k+1}(t)) - X(\pi_{k+1}(s))| &\leq \\ &\leq d(\pi_{k+1}(t), \pi_{k+1}(s))(3K^2 2^{k+1} + K^2 \psi^{-1}(\frac{1}{N_{k+1}^3} V_{k+1})),\end{aligned}$$

where

$$V_{k+1} = \sum_{u, v \in T_{k+1}} \varphi\left(\frac{|X(u) - X(v)|}{d(u, v)}\right).$$

Note that by the definition of  $k(s, t)$

$$d(\pi_{k+1}(t), \pi_{k+1}(s)) \leq d(t, T_{k+1}) + d(s, t) + d(s, T_{k+1}) \leq 2d(s, t).$$

Therefore using that  $\sum_{k=0}^{\infty} \frac{1}{N_{k+1}^3} V_{k+1} \leq Z$

$$|X(\pi_{k+1}(s)) - X(\pi_{k+1}(t))| \leq 6K^2 2^{k+1} d(s, t) + 2K^2 d(s, t) \varphi^{-1}(Z).$$

and hence

$$|X(\pi_{k+1}(s)) - X(\pi_{k+1}(t))| \leq 15K^2(2^{k+1} - 1)d(s, t) + 2K^2 d(s, t) \varphi^{-1}(Z). \quad (8)$$

Now we have to sum up bounds (6), (7), (8) and apply (5). The proof is completed with  $\bar{A} = 30K^2$  and  $\bar{B} = 10K^2$ . ■

**Corollary 2** *The following inequality holds*

$$\mathbf{E} \sup_{s, t \in T} \psi\left(\frac{|X(s) - X(t)| - A\tau(s, t)_+}{Bd(s, t)}\right) \leq 1,$$

where  $A, B$  are constant for Proposition 2.

The meaning of the result is that for exponentially concentrated random variables there is always some form Law of Iterated Logarithm. For example on the interval  $[0, 1] \subset \mathbb{R}$  and any fractional Brownian motion  $X(t)$ ,  $t \in T$  with the Hurst exponent  $H \in (0, 1)$ , we have  $d(s, t) = |s - t|^H$  and with usual  $T_n = \{kN_n^{-1} : k \in \{1, \dots, N_n\}\}$  we get  $\tau(s, t) \sim |s - t|^H \sqrt{\log_2(1 + |s - t|^{-H})}$ .

**Corollary 3** *If  $X(t)$ ,  $t \in [0, 1]$  is a fractional Brownian motion with the Hurst coefficient  $H \in (0, 2)$  then there exists universal  $A, B$  such that*

$$\mathbf{E} \sup_{s, t \in [0, 1]} \varphi_2\left(\frac{(|X(s) - X(t)| - A|s - t|^H \sqrt{\log_2(1 + |s - t|^{-H})})_+}{B|s - t|^H}\right) \leq 1.$$

## 4 The concentration inequality for two distances

As the basic example of the Bernstein inequality shows one distance may not suffice to fully describe the concentration property. The classical situation when this happens concerns independent, symmetric, identically distributed  $X_i$ ,  $1 \leq i \leq n$  of log concave tails, i.e. when  $u \rightarrow -\log \mathbf{P}(|X_i| > u)$  is a convex function for  $u \geq 0$ . We can consider canonical type process using  $T \subset \mathbb{R}^n$  and  $X(t) = \sum_{i=1}^n t_i X_i$ . Therefore we generalize slightly the idea described in the previous sections towards the case where two distances can be applied.

Assume that on set  $T$  there are two distances  $d_1, d_2$  that imply the same topology on  $T$  (i.e. we require that the convergence in  $d_1$  is equivalent to the convergence in  $d_2$ ). The main assumption on increments, is that there exists Young functions  $\psi_1$  and  $\psi_2$  such that

$$\mathbf{E} \min\left\{\psi_1\left(\frac{|X(s) - X(t)|}{d_1(s, t)}\right), \psi_2\left(\frac{|X(s) - X(t)|}{d_2(s, t)}\right)\right\} \leq 1 \quad \text{for } s, t \in T. \quad (9)$$

The same remark as for one distance control is valid. We can extend the approach on  $\psi_1, \psi_2$  that are increasing to infinity, continuous and  $\psi_i(0) = 0$ ,  $\psi_i(1) = 1$  for  $i = 1, 2$ . Moreover we assume the condition (3) for  $\psi_1$  and  $\psi_2$ , yet

it is not enough for the our analysis. We need a polynomial comparability of  $\psi_1^{-1}$  and  $\psi_2^{-1}$ . We state the condition in the general form but to avoid technical complications we assume that there exists a single  $\psi$  that satisfies (3) such that  $\psi_1(x) = \psi(x^{p_1})$ , and  $\psi_2(x) = \psi(x^{p_2})$  for  $p_1, p_2 > 0$ . The simplest case of the setting is when  $\psi_1 = \varphi_{p_1}$  and  $\psi_2 = \varphi_{p_2}$  for some  $p_1, p_2 \geq 1$  since then  $\psi_1$  and  $\psi_2$  are convex.

For simplicity let us assume that  $T$  is compact in the topology defined by  $d_1$  and  $d_2$ . Moreover we require that  $\psi$  such that  $\psi_1(x) = \psi(x^{p_1})$  and  $\psi_2(x) = \psi(x^{p_2})$  are convex.

The main point is that we define the common approximation net for both two distances. Let  $N_n = \psi(2^n)$ , for  $n \geq 1$  and  $N_0 = 1$ . Again we assume that  $(T_n)_{n=0}^\infty$  is admissible, i.e.  $|T_n| \leq N_n$ ,  $T_n \subset T_{n+1}$  and  $\bigcup_{n=0}^\infty T_n$  is dense in  $T$ . Consequently  $\psi_1^{-1}(N_n) = 2^{p_1 n}$  and  $\psi_2^{-1}(N_n) = 2^{p_2 n}$ . We extend our definition of  $\sigma_m(t)$ , i.e. we define

$$\sigma_m^j(t) = \sum_{n=m}^{\infty} 2^{p_j n} d_j(t, T_n), \quad j \in 1, 2.$$

W.l.o.g. we may assume that  $\sigma^j(t) < \infty$  for  $j = 1, 2$ . Moreover let

$$Z = \sum_{n=1}^{\infty} \frac{1}{N_n^3} \sum_{t \in T_n} \min\left\{\psi_1\left(\frac{|X(t) - X(\pi_{n-1}(t))|}{d_1(t, \pi_{n-1}(t))}\right), \psi_2\left(\frac{|X(t) - X(\pi_{n-1}(t))|}{d_2(t, \pi_{n-1}(t))}\right)\right\}.$$

Fix  $t \in T$ , we aim to extend Proposition 1. Obviously again for a given  $t \in T$  we can use the sequence  $n_i = n_i(t)$  such that  $n_0 = m$  and it increases to  $\infty$ , but the definition is more complicated. Let  $n_0 = m$  and

$$\begin{aligned} n_i &= \inf\{n > n_{i-1} : 2^{p_1 n} d_1(t, T_{n_{i-1}}) + 2^{p_2 n} d_2(t, T_{n_{i-1}}) > \\ &> 2(2^{p_1 n} d_1(t, T_n) + 2^{p_2 n} d_2(t, T_n))\}. \end{aligned}$$

In particular it means that either  $d_1(t, T_{n_{i-1}}) > 2d_1(t, T_{n_i})$  or  $d_2(t, T_{n_{i-1}}) > 2d_2(t, T_{n_i(t)})$ . However we cannot claim that the property holds for both two distances at once. It should be noticed also why  $n_i$  necessarily exists and this is due to the assumption  $\sigma_j(t) < \infty$  for  $j \in \{1, 2\}$ .

To formulate the best possible result we define for each  $t \in T$

$$\bar{d}_1(t, \pi_m(t)) = \sum_{i=0}^{\infty} d_1(t, T_{n_i}), \quad \bar{d}_2(t, \pi_m(t)) = \sum_{i=0}^{\infty} d_2(t, T_{n_i}).$$

Obviously in if  $T_n$  well approximates  $T$  in both two distances then  $\bar{d}_j(t, \pi_m(t))$  is comparable with  $d_j(t, \pi_m(t))$  but in full generality we cannot claim the statement. On the other hand there are estimates on  $\bar{d}_j(t, \pi_m(t))$  in terms of  $\sigma_m^j(t)$ , i.e.

$$\bar{d}_j(t, \pi_m(t)) \leq 2^{-p_j m} \sigma_m^j(t), \quad \text{for } j \in \{1, 2\}.$$

Usually the above bounds are sufficiently strong for concentration reason.

**Proposition 3** *Let  $X(t)$ ,  $t \in T$  satisfy (9) with  $\psi_1 = \psi(x^{p_1})$  and  $\psi_2 = \psi(x^{p_2})$  which verify 3 with constants  $K_1, K_2$  respectively. For all  $t \in T$  and  $m \geq 0$  the*



following inequality holds

$$|X(t) - X(\pi_m(t))| \leq A(\sigma_m^1(t) + \sigma_m^2(t)) + B(\bar{d}_1(t, \pi_m(t))\psi_1^{-1}(Z) + \bar{d}_2(t, \pi_m(t))\psi_2^{-1}(Z)),$$

where  $A = 3(1 + 2^{1+p})K^2$ ,  $B = 2K^2$  and  $K = \max\{K_1, K_2\}$ ,  $p = \max\{p_1, p_2\}$ .

**Proof.** We use the sequence  $n_i = n_i(t)$ ,  $i \geq 0$  to get

$$|X(t) - X(\pi_m(t))| \leq \sum_{i=0}^{\infty} |X(\pi_{n_i}(t)) - X(\pi_{n_{i+1}}(t))|. \quad (10)$$

Consequently due to (3)

$$\begin{aligned} & |X(\pi_{n_i}(t)) - X(\pi_{n_{i-1}}(t))| \leq \\ & \leq \sum_{j \in \{1,2\}} d_j(\pi_{n_i}(t), \pi_{n_{i-1}}(t))\psi_j^{-1}\left(\frac{N_{n_i}^3}{N_{n_i}^3} V_{n_i}\right) \leq \\ & \leq \sum_{j \in \{1,2\}} d_j(\pi_{n_i}(t), \pi_{n_{i-1}}(t)) [3K_j^2 2^{p_j n_i} + K_j^2 \psi_j^{-1}\left(\frac{1}{N_{n_i}^3} \sum_{u,v \in T_{n_i}} V_{n_i}\right)], \end{aligned}$$

where

$$V_n = \sum_{u,v \in T_n} \min\left\{\psi_1\left(\frac{|X(u) - X(v)|}{d_1(u,v)}\right), \psi_2\left(\frac{|X(u) - X(v)|}{d_2(u,v)}\right)\right\} \text{ for } n \geq 0.$$

Observe that for  $j \in \{1, 2\}$

$$d_j(\pi_{n_i}(t), \pi_{n_{i-1}}(t)) \leq d_j(t, T_{n_i}) + d_j(t, T_{n_{i-1}})$$

and therefore by the construction of  $n_i$

$$\begin{aligned} & \sum_{j \in \{1,2\}} 3K_j^2 d_j(\pi_{n_i}(t), \pi_{n_{i-1}}(t)) 2^{p_j n_i} \leq 3K^2 \sum_{j \in \{1,2\}} (d_j(t, T_{n_i}) + d(t, T_{n_{i-1}})) 2^{p_j n_i} \leq \\ & \leq 3K^2 \sum_{j \in \{1,2\}} (d(t, T_{n_i}) + 2d(t, T_{n_{i-1}})) 2^{p_j n_i}. \end{aligned}$$

Consequently

$$\sum_{i=0}^{\infty} \sum_{j \in \{1,2\}} 3K_j^2 d_j(\pi_{n_i}(t), \pi_{n_{i-1}}(t)) 2^{p_j n_i} \leq 3(1 + 2^{1+p})K^2(\sigma_m^1(t) + \sigma_m^2(t)).$$

On the other hand

$$\sum_{i=0}^{\infty} K_j^2 d_j(\pi_{n_i}(t), \pi_{n_{i-1}}(t)) \leq 2K_j^2 \bar{d}_j(t, \pi_m(t)).$$

and  $\sum_{i=0}^{\infty} \frac{1}{N_{n_i}^3} V_{n_i} \leq Z$ , hence

$$\begin{aligned} & |X(t) - X(\pi_m(t))| \leq 3(1 + 2^{1+p})K^2(\sigma_m^1(t) + \sigma_m^2(t)) + \\ & + 2K^2(\bar{d}_1(t, \pi_m(t))\psi_1^{-1}(Z) + \bar{d}_2(t, \pi_m(t))\psi_2^{-1}(Z)). \end{aligned}$$

It completes the proof with  $A = 3(1 + 2^{1+p})K^2$  and  $B = 2K^2$ , where  $K = \max\{K_1, K_2\}$  and  $p = \max\{p_1, p_2\}$ . ■

As for the modulus of continuity for given  $s, t \in T$  one has to be more careful. For the one distance control it was clear for which  $k \geq 0$  it is worth to jump from the approximation of  $t$  to the approximation of  $s$ . For the two distance control we have two possible solutions  $k^1(s, t)$  and  $k^2(s, t)$  such that for  $j \in \{1, 2\}$

$$k^j(s, t) = \max\{k \geq 0 : \sum_{x \in \{s, t\}} d_j(x, T_k) \geq d_j(s, t)\}.$$

Using that  $p_1, p_2 \geq 1$  can prove in the same way as for the one distance control that  $k^j(s, t)$ ,  $j \in \{1, 2\}$  which is the  $k \geq 0$  for which the function

$$\sigma_{k+1}^j(s) + \sigma_{k+1}^j(t) + \frac{2^{p_j(k+1)} - 1}{2^{p_j} - 1} d_j(s, t)$$

is the smallest possible. We use the same idea to define the general  $k(s, t)$  in the case we study.

First define  $k(s, t)$  as

$$k(s, t) = \max\{k \geq 0 : \sum_{j \in \{1, 2\}} \sum_{x \in \{s, t\}} 2^{p_j k} d_j(x, T_k) \geq \sum_{j \in \{1, 2\}} 2^{p_j k} d_j(s, t)\}.$$

As above we explain that for such  $k = k(s, t)$  the function

$$f(k) = \sum_{x \in \{s, t\}} \sum_{j \in \{1, 2\}} \sigma_{k+1}^j(x) + \sum_{j \in \{1, 2\}} \frac{2^{p_j(k+1)} - 1}{2^{p_j} - 1} d_j(s, t), \quad k \geq 0$$

is the smallest possible. Indeed observe that  $f(k) > f(k-1)$  implies that

$$\sum_{j \in \{1, 2\}} 2^{p_j k} d_j(s, t) \geq \sum_{j \in \{1, 2\}} \sum_{x \in \{s, t\}} 2^{p_j k} d_j(x, T_k),$$

which happens for all  $k > k(s, t)$ . On the other hand for all  $k \leq k(s, t)$  we have that  $f(k) \leq f(k-1)$ .

$$\sum_{j \in \{1, 2\}} 2^{p_j k} d_j(s, t) \geq \sum_{j \in \{1, 2\}} \sum_{x \in \{s, t\}} 2^{p_j k} d_j(x, T_k),$$

By the property of  $k^j(s, t)$ ,  $j \in \{1, 2\}$  we have mentioned above we get

$$\min\{k^1(s, t), k^2(s, t)\} \leq k(s, t) \leq \max\{k^1(s, t), k^2(s, t)\}. \quad (11)$$

We use  $k(s, t)$  to define the deterministic bound on  $|X(t) - X(s)|$ . Let

$$\bar{\tau}(s, t) = \sum_{x \in \{s, t\}} \sum_{j \in \{1, 2\}} \sigma_{k(s, t)+1}^j(x) + \sum_{j \in \{1, 2\}} \frac{2^{p_j(k(s, t)+1)} - 1}{2^{p_j} - 1} d_j(s, t).$$

The obvious result is

$$\tau_1(s, t) + \tau_2(s, t) \leq \bar{\tau}(s, t), \quad (12)$$

where

$$\tau_j(s, t) = \max\{\sigma_j(t, d(s, t)), \sigma_j(s, d(s, t))\}, \text{ for } s, t \in T, j \in \{1, 2\}$$

and

$$\sigma_j(t, a) = \sum_{n=0}^{\infty} 2^{p_j n} \min\{d_j(t, T_n), a\}, \quad a > 0, \quad t \in T.$$

Indeed for  $k = k(s, t)$  and  $j \in \{1, 2\}$

$$\begin{aligned} \tau_j(s, t) &\leq \max_{x \in \{s, t\}} \sigma_{k+1}^j(x) + \sum_{l=0}^k 2^{p_j l} d_j(s, t) \leq \\ &\leq \sigma_{k+1}^j(s) + \sigma_{k+1}^j(t) + \frac{2^{p_j(k+1)} - 1}{2^{p_j} - 1} d_j(s, t) \end{aligned}$$

and hence

$$\tau_1(s, t) + \tau_2(s, t) \leq \bar{\tau}(s, t).$$

Unfortunately  $\tau_1(s, t) + \tau_2(s, t)$  may be not comparable with  $\bar{\tau}(s, t)$ . Nevertheless we can still provide an upper bound, let  $k_1(s, t) = k_1$ ,  $k_2(s, t) = k_2$  the following holds

$$\begin{aligned} \bar{\tau}(s, t) &\leq 2(\tau_1(s, t) + \tau_2(s, t)) + \frac{2^{p_2}}{2^{p_2} - 1} (2^{k_1 p_2} - 2^{k_2 p_2}) d_2(s, t) 1_{k_1 \geq k_2} + \\ &+ \frac{2^{p_1}}{2^{p_1} - 1} (2^{k_2 p_1} - 2^{k_1 p_1}) d_1(s, t) 1_{k_2 \geq k_1}. \end{aligned} \quad (13)$$

Assume for simplicity that  $k_1 \leq k_2$ . By the definition of  $k(s, t)$

$$\begin{aligned} \bar{\tau}(s, t) &\leq \sigma_{k_2+1}^1(s) + \sigma_{k_2+1}^1(t) + \frac{2^{(k_2+1)p_1} - 1}{2^{p_1} - 1} d_1(s, t) + \\ &+ \sigma_{k_2+1}^2(s) + \sigma_{k_2+1}^2(t) + \frac{2^{(k_2+1)p_2} - 1}{2^{p_2} - 1} d_2(s, t). \end{aligned}$$

Now observe that

$$\sigma_{k_2+1}^2(s) + \sigma_{k_2+1}^2(t) + \frac{2^{(k_2+1)p_2} - 1}{2^{p_2} - 1} d_2(s, t) \leq 2\tau_2(s, t)$$

and using that  $k_1 \leq k_2$

$$\begin{aligned} \sigma_{k_2+1}^1(s) + \sigma_{k_2+1}^1(t) + \frac{2^{(k_2+1)p_1} - 1}{2^{p_1} - 1} d_1(s, t) &\leq \sigma_{k_1+1}^1(s) + \sigma_{k_1+1}^1(t) + \\ &+ \frac{2^{(k_1+1)p_1} - 1}{2^{p_1} - 1} d_1(s, t) + \frac{2^{(k_2+1)p_1} - 2^{(k_1+1)p_1}}{2^{p_1} - 1} d_1(s, t) \leq \\ &\leq 2\tau_1(s, t) + \frac{2^{p_2}}{2^{p_2} - 1} (2^{k_2 p_1} - 2^{k_1 p_1}) d_1(s, t). \end{aligned}$$

It implies (13). To understand this inequality better we have to observe that

$$2^{k_2 p_1} \leq (2^{k_2 p_2})^{\frac{p_1}{p_2}} \leq \left( \frac{\tau_2(s, t)}{d_2(s, t)} \right)^{\frac{p_1}{p_2}}$$

and hence

$$\frac{2^{p_1}}{2^{p_1}-1}(2^{k_2 p_1} - 2^{k_1 p_1})d_1(s, t) \leq \frac{2^{p_2}}{2^{p_1}-1} \left( \frac{\tau_2(s, t)}{d_2(s, t)} \right)^{\frac{p_1}{p_2}}.$$

Consequently

$$\begin{aligned} \bar{\tau}(s, t) &\leq \tau_1(s, t) + \tau_2(s, t) + \\ &+ \max\left\{ \frac{2^{p_2}}{2^{p_2}-1} \left( \frac{\tau_2(s, t)}{d_2(s, t)} \right)^{\frac{p_1}{p_2}} d_1(s, t), \frac{2^{p_2}}{2^{p_1}-1} \left( \frac{\tau_1(s, t)}{d_1(s, t)} \right)^{\frac{p_2}{p_1}} d_2(s, t) \right\} \end{aligned}$$

We need also a slight generalization of the distances  $d_1$  and  $d_2$ , namely

$$\bar{d}_j(s, t) = \sum_{i=0}^{\infty} d_j(t, T_{n_i(t)}) + d_j(s, t) + \sum_{i=0}^{\infty} d_j(s, T_{n_i(s)}),$$

and  $n_i(t), n_i(s)$  are sequences  $n_i$  for  $t, s$  (recall that  $n_i$  depend on points in  $T$ ) that starts from  $k(s, t) + 1$ . Again observe that

$$\bar{d}_1(s, t) + \bar{d}_2(s, t) \leq \max\{2^{-p_1 k(s, t)}, 2^{-p_2 k(s, t)}\} \bar{\tau}(s, t). \quad (14)$$

We may state the main result of these section.

**Proposition 4** *For each admissible net  $(T_n)_{n=0}^{\infty}$ , and separable  $X(t)$ ,  $t \in T$  that satisfies (1) the following inequality holds. For all  $s, t \in T$*

$$|X(s) - X(t)| \leq \bar{A}(\bar{\tau}(s, t)) + \bar{B}[\bar{d}_1(s, t)\psi_1^{-1}(Z) + \bar{d}_2(s, t)\psi_2^{-1}(Z)],$$

where  $\bar{A} = 3(1 + 2^{1+p})K^2$ ,  $\bar{B} = 5K^2$ ,  $Z \geq 0$  and  $\mathbf{E}Z \leq 1$ .

**Proof.** Let  $k = k(s, t)$  for  $s, t \in T$ . Observe that

$$\begin{aligned} |X(t) - X(s)| &\leq |X(t) - X(\pi_{k+1}(t))| + \\ &+ |X(s) - X(\pi_{k+1}(t))| + |X(\pi_{k+1}(t)) - X(\pi_{k+1}(s))|. \end{aligned} \quad (15)$$

By Proposition 3 we have for  $x \in \{s, t\}$

$$\begin{aligned} |X(x) - X(\pi_{k+1}(x))| &\leq A[\sigma_{k+1}^1(x) + \sigma_{k+1}^2(x)] + \\ &+ B[\bar{d}_1(x, \pi_k(x))\psi_1^{-1}(Z) + \bar{d}_2(x, \pi_k(x))\psi_2^{-1}(Z)], \end{aligned} \quad (16)$$

where  $A = 3(1 + 2^{1+p})K^2$  and  $B = 2K^2$ . We use (3) for  $\psi_1$  and  $\psi_2$  with constants  $K_1, K_2$

$$\begin{aligned} |X(\pi_{k+1}(t)) - X(\pi_{k+1}(s))| &\leq \\ &\leq \sum_{j \in \{1, 2\}} d_j(\pi_{k+1}(t), \pi_{k+1}(s))(3K_j^2 2^{p_j(k+1)} + K_j^2 \psi_j^{-1}(\frac{1}{N_{k+1}^3} V_{k+1})). \end{aligned}$$

where

$$V_{k+1} = \sum_{u, v \in T_{k+1}} \min\left\{ \psi_1\left(\frac{|X(u) - X(v)|}{d_1(u, v)}\right), \psi_2\left(\frac{|X(u) - x(v)|}{d_2(u, v)}\right) \right\}.$$

Note that

$$d_j(\pi_{k+1}(t), \pi_{k+1}(s)) \leq d_j(t, T_{k+1}) + d_j(s, t) + d_j(s, T_{k+1}).$$

Therefore by the construction of  $k$

$$\sum_{j \in \{1,2\}} K_j^2 2^{p_j(k+1)} d_j(\pi_{k+1}(s), \pi_{k+1}(t)) \leq 2K^2 \sum_{j \in \{1,2\}} 2^{p_j(k+1)} d_j(s, t).$$

Consequently

$$\begin{aligned} |X(\pi_{k+1}(s)) - X(\pi_{k+1}(t))| &\leq 6K^2 \sum_{j \in \{1,2\}} 2^{p_j(k+1)} d_j(s, t) + \\ &+ K^2 [\bar{d}_1(s, t) \psi_1^{-1}(Z) + \bar{d}_2(s, t) \psi_2^{-1}(Z)]. \end{aligned}$$

Therefore

$$\begin{aligned} |X(\pi_{k+1}(s)) - X(\pi_{k+1}(t))| &\leq 6K^2 2^p \sum_{j \in \{1,2\}} \frac{2^{p_j(k+1)} - 1}{2^{p_j} - 1} d_j(s, t) + \\ &+ K^2 [\bar{d}_1(s, t) \psi_1^{-1}(Z) + \bar{d}_2(s, t) \psi_2^{-1}(Z)]. \end{aligned} \quad (17)$$

Now we have to sum up bounds (15), (16), (17)

$$|X(s) - X(t)| \leq 2A\tau(s, t) + 5K^2 [\bar{d}_1(s, t) \psi_1^{-1}(Z) + \bar{d}_2(s, t) \psi_2^{-1}(Z)].$$

The proof is completed with  $\bar{A} = 6(1 + 2^{1+p})K^2$  and  $B = 5K^2$ . ■

The consequence of the theorem is the Bernstein type inequality for chaining. Using (14) and Proposition 3.

**Corollary 4** *The following inequality holds*

$$\mathbf{E} \min\{\psi_1\left(\frac{|X(s) - X(t)| - \bar{A}\tau(s, t)}{\bar{B}\bar{d}_1(s, t)}\right), \psi_2\left(\frac{|X(s) - X(t)| - \bar{A}\tau(s, t)}{\bar{B}\bar{d}_2(s, t)}\right)\} \leq 1.$$

Moreover

$$\bar{d}_j(s, t) \leq 2^{-p_j(k(s,t)+1)} \tau_j(s, t), \quad \text{for } j \in \{1, 2\},$$

where  $k(s, t)$  is defined by (14).

Obviously we natural application of the idea should be to the theory of empirical processes. We discuss the question for a certain question in the following section.

## 5 Square estimation for one distance

The following problem was studied in [2]. Let  $X_1, X_2, \dots, X_N$  be independent random variables with values in a measurable space  $(\mathcal{X}, \mathcal{B})$ . Let  $\mathcal{F}$  be a family of real measurable functions on  $(\mathcal{X}, \mathcal{B})$ . To explain the problem of square estimation we start from the analysis of one distance control and then we turn to consider more complicated case of two distance control.

Let us assume that there exists distance  $d$  on  $\mathcal{F}$  such that for any given  $f, g \in \mathcal{F}$

$$\mathbf{P}(|(f - g)(X_i)| \geq u) \leq 2 \exp(-\frac{u^2}{d(f, g)^2}), \quad (18)$$

for all  $1 \leq i \leq N$ ,  $u > 0$ . Moreover we assume that

$$\mathbf{P}(|f(X_i)| \geq u) \leq 2 \exp(-\frac{u^2}{d(f, 0)^2}) \quad (19)$$

Recall that it is equivalent to  $\|(f - g)(X_i)\|_{\varphi_2} \leq Cd(f, g)$  and  $\|f(X_i)\|_{\varphi_2} \leq Cd(f, 0)$  for all  $i \in \{1, 2, \dots, N\}$  and a universal constant  $C > 0$ . Obviously in particular

$$\mathbf{E}(f - g)^2(X_i) \leq 2C^2 d(f, g)^2 \quad \text{and} \quad \frac{1}{N} \mathbf{E} \sum_{i=1}^N (f - g)^2(X_i) \leq 2C^2 d(f, g)^2 \quad (20)$$

and similarly

$$\mathbf{E}f^2(X_i) \leq 2C^2 d(f, 0)^2 \quad \text{and} \quad \frac{1}{N} \mathbf{E} \sum_{i=1}^N f^2(X_i) \leq 2C^2 d(f, 0)^2 \quad (21)$$

We aim to provide a concentration inequality for

$$S_N(f) = \frac{1}{N} \sum_{i=1}^N (f^2 - \mathbf{E}f^2)(X_i).$$

To state the result we have to recall Bernstein type inequalities. For all  $f, g \in \mathcal{F}$  the following holds

$$\mathbf{P}(|S_N(f - g)| \geq u) \leq 2 \exp(-N \min(\frac{u^2}{4d(f, g)^4}, \frac{u}{4d(f, g)^2})) \quad (22)$$

Moreover for all  $f \in \mathcal{F}$

$$\mathbf{P}(|S_N(f)| \geq u) \leq 2 \exp(-N \min(\frac{u^2}{4d(f, 0)^4}, \frac{u}{4d(f, 0)^2})).$$

The above inequalities can be rewritten using the following function

$$\varphi(x) = 2^{\min\{\sqrt{N}x, x^2\}} - 1, \quad \varphi^{-1}(x) = \max\{\frac{1}{\sqrt{N}} \log_2(1 + x), \sqrt{\log_2(1 + x)}\}.$$

The function  $\varphi$  is not convex but is comparable to a convex function say  $\psi(x) = 2^{\min\{2\sqrt{N}x - N, x^2\}}$  (note that  $\psi'(\sqrt{N}) = 2 \log 2 \sqrt{N} 2^N$ ). Clearly

$$\psi(x/2) \leq \varphi(x) \leq \psi(x), \quad \text{for all } x > 0.$$

As usual  $\varphi(0) = 0$  and  $\varphi(1) = 1$  and  $\varphi$  satisfies (3) with  $K = 1$ . Moreover  $\varphi(D^{-1}x) \leq D^{-1}\varphi(x)$  for  $D \geq 1$  and  $x \geq 0$ . The meaning of (18) in terms of  $\varphi$  is that  $f, g \in \mathcal{F}$

$$\|S_N(f - g)\|_{\varphi} \leq K^2 N^{-\frac{1}{2}} d(f, g)^2. \quad (23)$$

In the same way the meaning of (19) is such that for each  $f \in \mathcal{F}$

$$\|S_N(f)\|_\varphi \leq K^2 N^{-\frac{1}{2}} d(f, 0)^2. \quad (24)$$

On the other hand

$$\|f^2 - g^2\|_{\varphi_1} \leq \|f - g\|_{\varphi_2} \|f + g\|_{\varphi_2} \leq C^2(d(f, 0) + d(g, 0))d(f, g).$$

Consequently for any  $f, g \in \mathcal{F}$

$$\|S_N(f) - S_N(g)\|_\varphi \leq K^2 N^{-\frac{1}{2}} (d(f, 0) + d(g, 0))d(f, g). \quad (25)$$

We are ready to state the main result of the section.

**Theorem 1** *Suppose that  $\alpha = \sup_{t \in \mathcal{F}} d(f, 0) < \infty$  for all  $f \in \mathcal{F}$  then*

$$\sup_{f \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^N (f^2 - \mathbf{E}f^2)(X_i) \leq A(N^{-1}\gamma_2^2(\mathcal{F}) + N^{-\frac{1}{2}}\alpha\gamma_2(\mathcal{F})) + BN^{-\frac{1}{2}}\alpha^2\varphi^{-1}(Z),$$

where  $Z \geq 0$  is such that  $\mathbf{E}Z \leq 1$  and  $A, B$  are universal constants. In terms of concentration it means that for all  $u \geq 0$

$$\begin{aligned} & \mathbf{P}\left(\sup_{f \in \mathcal{F}} \frac{1}{N} \left(\sum_{i=1}^N (f^2 - \mathbf{E}f^2)(X_i)\right) > \right. \\ & \left. > \bar{A}(N^{-1}\gamma_2^2(\mathcal{F}) + N^{-\frac{1}{2}}\alpha\gamma_2(\mathcal{F})) + \bar{B}N^{-\frac{1}{2}}\alpha^2u\right) \leq \exp(-\min(\sqrt{Nu}, u^2)), \end{aligned}$$

where  $\bar{A}, \bar{B}$  are universal constants.

**Proof.** We mimic the proof of Proposition 1. Assume there is an admissible sequence of nets  $(\mathcal{F}_n)_{n=0}^\infty$  such that  $\mathcal{F}_n \subset \mathcal{F}_{n+1}$  and  $|\mathcal{F}_n| \leq N_n = \varphi_2(2^n) = 2^{2^n} - 1$  for  $n \geq 1$  and  $|\mathcal{F}_0| = 1$  and  $\bigcup_{n=0}^\infty \mathcal{F}_n$  is dense in  $\mathcal{F}$ . There exists  $k_0 \geq 0$  such that  $2^{2^{k_0}} \leq N < 2^{2^{(k_0+1)}}$  and hence

$$\varphi^{-1}(N_n) = \begin{cases} \frac{1}{\sqrt{N}} 2^{2^n} & n > k_0 \\ 2^n & n \leq k_0 \end{cases}$$

Therefore  $\varphi^{-1}(N_n)$  behaves like  $\varphi_2^{-1}(N_n)$  for  $n \leq k_0$  and  $N^{-\frac{1}{2}}\varphi_1^{-1}(N_n)$  for  $n > k_0$ . Fix  $f \in \mathcal{F}$ . The main idea is to apply the following split

$$\begin{aligned} f^2(X_i) &= (f - \pi_{k_0}(f) + \pi_{k_0}(f))^2(X_i) = (f(X_i) - \pi_{k_0}(f))^2 + \\ &+ 2(f - \pi_{k_0}(f))\pi_{k_0}(f)(X_i) + (\pi_{k_0}(f))^2(X_i). \end{aligned}$$

which implies that

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N f^2(X_i) &\leq \frac{1}{N} \sum_{i=1}^N (f - \pi_{k_0}(f))^2(X_i) + \\ &+ 2\frac{1}{N} \sum_{i=1}^N (f - \pi_{k_0}(f))\pi_{k_0}(f)(X_i) + \frac{1}{N} \sum_{i=1}^N (\pi_{k_0}(f))^2(X_i). \end{aligned}$$

Let us define

$$\begin{aligned} P_N(f) &= \frac{1}{N} \sum_{i=1}^N (f - \pi_{k_0}(f))^2(X_i), \\ Q_N(f) &= \frac{2}{N} \sum_{i=1}^N ((f - \pi_{k_0}(f))\pi_{k_0}(f))(X_i), \\ R_N(f) &= \frac{1}{N} \sum_{i=1}^N (\pi_{k_0}(f))^2(X_i). \end{aligned}$$

Consequently

$$S_N(f) = P_N(f) + Q_N(f) + R_N(f) - \mathbf{E}P_N(f) - \mathbf{E}Q_N(f) - \mathbf{E}R_N(f).$$

Now we aim to study  $P_N(f)$ ,  $Q_N(f)$  and  $R_N(f)$  separately.

We start from  $P_N(f)$ . The main point is to explore our concept of certain structure  $n_k$ ,  $k \geq 0$  that may depend on  $f$  to suitably bound  $P_N(f)$ . We use the following chaining

$$(f - \pi_{k_0}(f))(X_i) = \sum_{k=0}^{\infty} (\pi_{n_k}(f) - \pi_{n_{k+1}}(f))(X_i),$$

where  $(n_k)_{k=0}^{\infty}$  is defined as follows  $n_0(f) = k_0$  and

$$n_i = \inf\{n > n_{i-1} : 2d(f, \mathcal{F}_n) < d(f, \mathcal{F}_{n_{i-1}})\}.$$

Let  $\Delta_k(f) = \frac{1}{N} \sum_{i=1}^N (\pi_{n_k}(f) - \pi_{n_{k+1}}(f))^2(X_i)$ . By the triangle inequality it is clear that

$$\left[\frac{1}{N} \sum_{i=1}^N (f - \pi_{k_0}(f))^2(X_i)\right]^{\frac{1}{2}} \leq \sum_{k=0}^{\infty} [\Delta_k(f)]^{\frac{1}{2}}. \quad (26)$$

Moreover

$$[\Delta_k(f)]^{\frac{1}{2}} \leq [(\Delta_k(f) - \mathbf{E}\Delta_k(f))_+]^{\frac{1}{2}} + [\mathbf{E}\Delta_k(f)]^{\frac{1}{2}}.$$

Clearly

$$[\mathbf{E}\Delta_k(f)]^{\frac{1}{2}} \leq Cd(\pi_{n_k}(f), \pi_{n_{k+1}}(f)). \quad (27)$$

By the Bernstein inequality

$$\|(\Delta_k(f) - \mathbf{E}\Delta_k(f))_+\|_{\varphi} \leq K^2 N^{-\frac{1}{2}} d(\pi_{n_k}(f), \pi_{n_{k+1}}(f))^2.$$

Therefore

$$\begin{aligned} & [(\Delta_k(f) - \mathbf{E}\Delta_k(f))_+]^{\frac{1}{2}} \leq \\ & \leq KN^{-\frac{1}{4}} d(\pi_{n_k}(f), \pi_{n_{k+1}}(f)) (3\varphi^{-1}(N_{k+1}) + \varphi^{-1}(Z_1))^{\frac{1}{2}} \leq \\ & \leq KN^{-\frac{1}{4}} d(\pi_{n_k}(f), \pi_{n_{k+1}}(f)) (2[\varphi^{-1}(N_{k+1})]^{\frac{1}{2}} + [\varphi^{-1}(Z_1)]^{\frac{1}{2}}), \end{aligned} \quad (28)$$

where  $Z_1$  is given by

$$Z_1 = \sum_{k=1}^{\infty} \frac{1}{N_k^3} U_k, \quad U_k = \sum_{g, h \in \mathcal{F}_k} \varphi(X(g, h))$$



and

$$X(g, h) = \frac{\sum_{i=1}^N [(g - h)^2 - \mathbf{E}(g - h)](X_i)}{K^2 N^{\frac{1}{2}} d(g, h)^2}.$$

Clearly  $\mathbf{E}Z_1 \leq 1$  by (25). We can sum up the bounds, i.e. using (26), (27) and (28) we get

$$\begin{aligned} [P_N(f)]^{\frac{1}{2}} &\leq \sum_{k=0}^{\infty} d(\pi_{n_k}(f), \pi_{n_{k+1}}(f))(C + 2KN^{-\frac{1}{4}}[\varphi^{-1}(N_{k+1})]^{\frac{1}{2}}) + \\ &+ KN^{-\frac{1}{4}} \sum_{k=0}^{\infty} d(\pi_{n_k}(f), \pi_{n_{k+1}}(f))[\varphi^{-1}(Z_1)]^{\frac{1}{2}}. \end{aligned}$$

We have noticed that  $\varphi^{-1}(N_n) = N^{-\frac{1}{2}}2^{2n}$ , and by the construction

$$d(\pi_{n_{k+1}}(f), \pi_{n_k}(f)) \leq d(f, \pi_{n_k}(f)) + 2d(f, \pi_{n_{k-1}}(f))$$

and for  $k \geq 0$

$$d(\pi_{n_k}(f), f) \geq 2d(\pi_{n_{k+1}}(f), f).$$

It implies that

$$\begin{aligned} [P_N(f)]^{\frac{1}{2}} &\leq 12KN^{-\frac{1}{2}}\sigma_{k_0}(f) + 6Cd(f, \pi_{k_0}(f)) + \\ &+ 6CN^{-\frac{1}{4}}d(f, \pi_{k_0}(f))\varphi^{-1}(Z_1)^{\frac{1}{2}}, \end{aligned}$$

where  $\sigma_{k_0}(f) = \sum_{n=k_0}^{\infty} 2^n d(f, \mathcal{F}_n)$ . Obviously

$$\sup_{f \in \mathcal{F}} \sigma_{k_0}(f) \leq \gamma_2(\mathcal{F}).$$

Moreover

$$d(f, \pi_{k_0}(f)) \leq d(f, \pi_{k_0}(f))(2^{k_0+1}N^{-\frac{1}{2}}) \leq 2N^{-\frac{1}{2}}\gamma_2(\mathcal{F}) \quad (29)$$

and on the other hand

$$d(f, \pi_{k_0}(f)) \leq 2\alpha. \quad (30)$$

Since  $\bigcup_{n=0}^{\infty} \mathcal{F}_n$  is dense in  $\mathcal{F}$  we acquire for each  $f \in \mathcal{F}$

$$\begin{aligned} P_N(f) &\leq [12(K + C)N^{-\frac{1}{2}}\gamma_2(\mathcal{F}) + \\ &+ 12CN^{-\frac{1}{4}}\min\{\alpha, N^{-\frac{1}{4}}\gamma_2(\mathcal{F})\}[\varphi^{-1}(Z_1)]^{\frac{1}{2}}]^2 \leq 2(12)^2(K + C)^2N^{-1}\gamma_2^2(\mathcal{F}) + \\ &+ 2(12)^2C^2N^{-\frac{1}{2}}\min\{\alpha^2, N^{-\frac{1}{2}}\gamma_2^2(\mathcal{F})\}\varphi^{-1}(Z_1). \end{aligned} \quad (31)$$

Moreover by (29) and (30)

$$\mathbf{E}P_N(f) \leq C^2d(f, \pi_{k_0}(f))^2 \leq 4C^2\min\{\alpha^2, N^{-1}\gamma_2^2(\mathcal{F})\}. \quad (32)$$

The second point is to consider  $R_N(f)$ . We use the following chaining

$$\begin{aligned} &\frac{1}{N} \sum_{i=1}^N ((\pi_{k_0}(f))^2 - \mathbf{E}(\pi_{k_0}(f))^2 - (\pi_0(f))^2 + \mathbf{E}(\pi_0(f))^2)(X_i) \leq \\ &\leq \sum_{k=1}^{\infty} \frac{1}{N} \sum_{i=1}^N ((\pi_{n_k}(f))^2 - (\pi_{n_{k-1}}(f))^2 - \mathbf{E}(\pi_{n_k}(f))^2 + \mathbf{E}(\pi_{n_{k-1}}(f))^2)(X_i). \end{aligned}$$

where

$$n_i(f) = \min\{\inf\{n > n_{i-1}(f) : 2d(f, \mathcal{F}_n) < d(f, \mathcal{F}_{n_{i-1}})\}, k_0\}.$$

Note that  $n_i = k_0$  for a finite  $i \geq 0$ . Using (29) we get

$$\begin{aligned} & \left| \frac{1}{N} \sum_{i=1}^N ((\pi_{n_k}(f))^2 - (\pi_{n_{k-1}}(f))^2 - \mathbf{E}(\pi_{n_k}(f))^2 + \mathbf{E}(\pi_{n_{k-1}}(f))^2)(X_i) \right| \leq \\ & \leq 2\alpha N^{-\frac{1}{2}} K^2 d(\pi_{n_k}(f), \pi_{n_{k-1}}(f)) (3\varphi^{-1}(N_{n_k}) + \varphi^{-1}(Z_2)), \end{aligned}$$

where

$$Z_2 = \sum_{k=1}^{\infty} \frac{1}{N_k^3} V_k, \quad V_k = \sum_{g, h \in \mathcal{F}_k} \varphi(X(g, h)),$$

and

$$X(g, h) = \frac{\sum_{i=1}^N (g^2 - h^2 - \mathbf{E}g^2 + \mathbf{E}h^2)}{2K^2 N^{\frac{1}{2}} \alpha d(g, h)}.$$

Note that  $\mathbf{E}Z_2$  due to (23). By the construction of  $n_k$ ,  $k \geq 1$

$$d(\pi_{n_k}(f), \pi_{n_{k-1}}(f)) \leq d(f, \pi_{n_k}(f)) + 2d(f, \pi_{n_{k-1}}(f))$$

and

$$d(f, \pi_{n_k}(f)) \geq 2d(f, \pi_{n_{k-1}}(f)).$$

Since  $\varphi^{-1}(N_n) = 2^n$  we obtain that

$$\begin{aligned} & \frac{1}{N} \sum_{i=1}^N ((\pi_{k_0}(f))^2 - \mathbf{E}(\pi_{k_0}(f))^2 - (\pi_0(f))^2 + \mathbf{E}(\pi_0(f))^2)(X_i) \leq \\ & \leq 12K^2 N^{-\frac{1}{2}} \alpha \sigma_{0, k_0}(f) + 6N^{-\frac{1}{2}} K^2 \alpha d(f, \pi_0(f)) \varphi^{-1}(Z_2), \end{aligned}$$

where

$$\sigma_{0, k_0}(f) = \sum_{n=0}^{k_0} 2^n d(f, \mathcal{F}_n) \leq \gamma_2(\mathcal{F}).$$

Obviously  $\mathbf{E}Z_2 \leq 1$ . On the other hand

$$\frac{1}{N} \sum_{i=1}^N ((\pi_0(f))^2 - \mathbf{E}(\pi_0(f))^2)(X_i) \leq K^2 N^{-\frac{1}{2}} \alpha^2 \varphi^{-1}(Z_3)$$

where

$$Z_3 = \varphi\left(\frac{\sum_{i=1}^N ((\pi_0(f))^2 - \mathbf{E}(\pi_0(f))^2)(X_i)}{K^2 N^{\frac{1}{2}} \alpha^2}\right).$$

Again observe that  $\mathbf{E}Z_3$  by (24). It implies that

$$R_N(f) - \mathbf{E}R_N(f) \leq 12K^2 N^{-\frac{1}{2}} \alpha \gamma_2(\mathcal{F}) + 6K^2 N^{-\frac{1}{2}} \alpha^2 (\varphi^{-1}(Z_2) + \varphi^{-1}(Z_3)). \quad (33)$$

Moreover

$$\mathbf{E}R_N(f) \leq C^2 d^2(\pi_{k_0}(f), f) \leq 4C^2 \min\{\alpha^2, N^{-1} \gamma_2^2(\mathcal{F})\}. \quad (34)$$

It remains to bound  $Q_N(f)$ . Clearly

$$Q_N(f) \leq 2[P_N(f)]^{\frac{1}{2}}[R_N(f)]^{\frac{1}{2}}. \quad (35)$$

and

$$[R_N(f)]^{\frac{1}{2}} \leq [(R_N(f) - \mathbf{E}R_N(f))_+]^{\frac{1}{2}} + [\mathbf{E}R_N(f)]^{\frac{1}{2}}.$$

Therefore using (34) and (35)

$$\begin{aligned} Q_N(f) &\leq 2([\mathbf{E}R_N(f)]^{\frac{1}{2}} + [(R_N(f) - \mathbf{E}R_N(f))_+]^{\frac{1}{2}})[P_N(f)]^{\frac{1}{2}} \leq \\ &\leq 2(2C \min\{\alpha, N^{-\frac{1}{2}}\gamma_2(\mathcal{F})\} + [(R_N(f) - \mathbf{E}R_N(f))_+]^{\frac{1}{2}})[P_N(f)]^{\frac{1}{2}}. \end{aligned}$$

and consequently

$$\begin{aligned} Q_N(f) &\leq 4C \min\{\alpha, N^{-\frac{1}{2}}\gamma_2(\mathcal{F})\}[P_N(f)]^{\frac{1}{2}} + \\ &+ 2(R_N(f) - \mathbf{E}R_N(f) + P_N(f)). \end{aligned} \quad (36)$$

Moreover by (34) and (32)

$$|\mathbf{E}Q_N(f)| \leq 2[\mathbf{E}P_N(f)]^{\frac{1}{2}}[\mathbf{E}R_N(f)]^{\frac{1}{2}} \leq 8C^2 \min\{\alpha^2, N^{-1}\gamma_2^2(\mathcal{F})\}. \quad (37)$$

Finally we can sum up the inequalities. First observe that

$$\begin{aligned} P_N(f) + Q_N(f) + R_N(f) - \mathbf{E}(P_N(f) + Q_N(f) + R_N(f)) &\leq P_N(f) + R_N(f) - \\ &\leq -\mathbf{E}R_N(f) + 2([(R_N(f) - \mathbf{E}R_N(f))_+]^{\frac{1}{2}} + [\mathbf{E}R_N(f)]^{\frac{1}{2}})[P_N(f)]^{\frac{1}{2}} + |\mathbf{E}Q_N(f)| \leq \\ &\leq 2P_N(f) + 2(R_N(f) - \mathbf{E}R_N(f)) + 2[\mathbf{E}R_N(f)]^{\frac{1}{2}}[P_N(f)]^{\frac{1}{2}} + |\mathbf{E}Q_N(f)| \end{aligned}$$

and hence using (31), (33) and (37)

$$\begin{aligned} P_N(f) + Q_N(f) + R_N(f) - \mathbf{E}(P_N(f) + Q_N(f) + R_N(f)) &\leq \\ &\leq 2[12(K+C)N^{-\frac{1}{2}}\gamma_2(\mathcal{F}) + 12CN^{-\frac{1}{4}}\min\{\alpha, N^{-\frac{1}{4}}\gamma_2(\mathcal{F})\}[\varphi^{-1}(Z_1)]^{\frac{1}{2}}]^2 + \\ &+ 2[12K^2N^{-\frac{1}{2}}\alpha\gamma_2(\mathcal{F}) + K^2N^{-\frac{1}{2}}\alpha^2(6\varphi^{-1}(Z_2) + 2\varphi^{-1}(Z_3))] + \\ &+ 4C \min\{\alpha, N^{-\frac{1}{2}}\gamma_2(\mathcal{F})\}[12(K+C)N^{-\frac{1}{2}}\gamma_2(\mathcal{F}) + \\ &+ 12CN^{-\frac{1}{4}}\min\{\alpha, N^{-\frac{1}{4}}\gamma_2(\mathcal{F})\}[\varphi^{-1}(Z_1)]^{\frac{1}{2}}] + \\ &+ 8C^2 \min\{\alpha^2, N^{-1}\gamma_2^2(\mathcal{F})\}. \end{aligned}$$

Therefore

$$\begin{aligned} P_N(f) + Q_N(f) + R_N(f) - \mathbf{E}(P_N(f) + Q_N(f) + R_N(f)) &\leq \\ &\leq (24)^2(K+C)^2N^{-1}\gamma_2^2(\mathcal{F}) + 24C(K+C)\min\{\alpha, N^{-\frac{1}{2}}\gamma_2(\mathcal{F})\}N^{-\frac{1}{2}}\gamma_2(\mathcal{F}) + \\ &+ 24K^2N^{-\frac{1}{2}}\alpha\gamma_2(\mathcal{F}) + 8C^2 \min\{\alpha^2, N^{-1}\gamma_2^2(\mathcal{F})\} + \\ &+ (24)^2C^2N^{-\frac{1}{2}}\min\{\alpha^2, N^{-\frac{1}{2}}\gamma_2^2(\mathcal{F})\}\varphi^{-1}(Z_1) + \\ &+ 6K^2N^{-\frac{1}{2}}\alpha^2(\varphi^{-1}(Z_2) + \varphi^{-1}(Z_3)). \end{aligned}$$

It proves that

$$\begin{aligned} P_N(f) + Q_N(f) + R_N(f) - \mathbf{E}(P_N(f) + Q_N(f) + R_N(f)) &\leq \\ &\leq A(N^{-1}\gamma_2^2(\mathcal{F}) + N^{-\frac{1}{2}}\alpha\gamma_2(\mathcal{F})) + BN^{-\frac{1}{2}}\varphi^{-1}(Z), \end{aligned}$$

where  $A = (24)^2 K^2 + 24C(K+C) + 8C^2 \leq (30)^2(K+C)$  and  $B = 6(24)^2 \max\{C^2, K^2\}$  and

$$Z = \frac{1}{2}Z_1 + \frac{1}{4}Z_2 + \frac{1}{4}Z_3.$$

Not that using that  $D^{-1}\varphi(x) \geq \varphi(D^{-1}x)$  for  $D \geq 1$

$$\begin{aligned} \frac{1}{2}\varphi^{-1}(Z_1) + \frac{1}{4}\varphi^{-1}(Z_2) + \frac{1}{4}\varphi^{-1}(Z_3) &\leq \varphi^{-1}\left(\frac{1}{2}Z_1\right) + \varphi^{-1}\left(\frac{1}{4}Z_2\right) + \varphi^{-1}\left(\frac{1}{4}Z_3\right) \leq \\ &\leq 3\varphi^{-1}\left(\frac{1}{2}Z_1 + \frac{1}{4}Z_2 + \frac{1}{4}Z_3\right) = 3\varphi^{-1}(Z). \end{aligned}$$

Clearly  $\mathbf{E}Z \leq 1$  which ends the proof. ■

## 6 Compressed sensing

In the compressed sensing we consider  $N \times M$  matrix  $A$ , where  $N \ll M$ . We want to reconstruct all vectors  $x \in \mathbb{R}^M$  of sparse support, i.e. all vectors such that  $|\{i \in \{1, \dots, M\} : x_i \neq 0\}| \leq m$ . For simplicity denote by  $\Sigma_m$  the space of  $m$ -sparse vectors in  $\mathbb{R}^M$ . The main tool to recover any sparse vector  $x$  is the  $l^1$  minimization which works [1] whenever  $A$  satisfies Restricted Isometry Property, namely

$$1 - \delta_m \leq |Ax|_2^2 \leq 1 + \delta_m$$

for all  $m$ -sparse  $x$  in the unit sphere of  $\mathbb{R}^M$  (which we denote by  $S_2(\Sigma_m)$ ). The main result for RIP is the following

**Theorem 2** *Let  $1 \leq m \leq N/2$ . Let  $A$  be an  $N \times M$  matrix. If  $\delta_{2m}(A) < \sqrt{2} - 1$ , then  $A$  satisfies the exact reconstruction property of order  $m$  (all vectors that are  $m$ -sparse are reconstructed).*

Consider  $A = N^{-\frac{1}{2}}(Y_1, \dots, Y_N)^T$  and let  $Y_i$  be independent subgaussians with  $\|Y_i\|_{\varphi_2} \leq \alpha$ . It is clear then

$$(1 - \delta_{2m})|x|_2^2 \leq |Ax|_2^2 \leq (1 + \delta_{2m})|x|_2^2,$$

where

$$\delta_{2m} = \sup_{x \in S_2(\Sigma_{2m})} \left| \frac{1}{N} \sum_{i=1}^N \langle Y_i, x \rangle^2 - 1 \right|.$$

To ensure  $\delta_{2m} < \sqrt{2} - 1$  we need that

$$\sup_{x \in S_2(\Sigma_{2m})} \left| \frac{1}{N} \sum_{i=1}^N \langle Y_i, x \rangle^2 - 1 \right| < \sqrt{2} - 1.$$

Note that for any  $x, y \in S_2$  by the Schwarz inequality

$$\|\langle Y_i, x - y \rangle\|_{\varphi_2} \leq \|Y_i\|_{\varphi_2} \|x - y\|_2 \leq \alpha \|x - y\|_2.$$

Let  $\mathcal{F} = \{f(\cdot) = \langle x, \cdot \rangle : x \in S_2(\Sigma_{2m})\}$ . Therefore for  $f(\cdot) = \langle x, \cdot \rangle$  and  $g(\cdot) = \langle y, \cdot \rangle$

$$d(f, g) \leq \alpha \|x - y\|_2, \quad f, g \in \mathcal{F}. \quad (38)$$

We can use the result of the previous section provided we can compute  $\gamma_2(S_2(\Sigma_{2m}))$ . By the majorizing measure theorem

$$\gamma_2(S_2(\Sigma_{2m})) \sim \mathbf{E}(\sup_{t \in S_2(\Sigma_{2m})} \sum_{i=1}^M t_i g_i),$$

where  $g_i$ ,  $1 \leq i \leq M$  are independent standard Gaussian variables. Moreover

$$\mathbf{E}(\sup_{t \in S_2(\Sigma_{2m})} \sum_{i=1}^M t_i g_i) = \mathbf{E}(\sum_{i=1}^{2m} (g_i^*)^2)^{\frac{1}{2}},$$

where  $g_i^*$  is a non-decreasing rearrangement of  $(|g_i|)_{i=1}^N$ . Finally we have a simple result to compute the last quantity.

**Lemma 1** *Let There exists absolute constants  $c_0, c_1 > 0$  such that the following holds. Let  $(g_i)_{i=1}^M$  be a family of independent standard Gaussian random variables and  $(g_i^*)_{i=1}^M$  its non-decreasing rearrangement. For any  $k \leq M$  we have*

$$\sqrt{c_1 k \log(k^{-1}M)} \leq \mathbf{E}(\sum_{i=1}^k (g_i^*)^2)^{\frac{1}{2}} \leq 2\sqrt{k \log(c_0 k^{-1}M)}.$$

It shows that

$$\gamma_2(S_2(\Sigma_{2m})) \leq K_0 \sqrt{m \log(c_0 m^{-1}M)},$$

where  $K_0, c_0$  are absolute constants. Therefore due to (38)

$$\gamma_2(\mathcal{F}) \leq K_0 \alpha \sqrt{m \log(c_0 m^{-1}M)}.$$

Denote  $\gamma = K_0(m \log(c_0 m^{-1}M))^{\frac{1}{2}}$ , by Theorem 1 we get

$$\begin{aligned} \mathbf{P}(\sup_{x \in S_2(\Sigma_{2m})} |\frac{1}{N} \sum_{i=1}^N \langle Y_i, x \rangle^2 - 1| > A(N^{-1}\alpha^2\gamma^2 + N^{-\frac{1}{2}}\alpha^2\gamma) + BN^{-\frac{1}{2}}\alpha^2u) \leq \\ \leq \exp(-(\log 2) \min\{N^{-\frac{1}{2}}u, u^2\}), \end{aligned}$$

where  $1 \leq A, B < \infty$ . Now it is clear that we should take  $u = c_0\gamma$  and then for any

$$(A + B + c_0)N^{-\frac{1}{2}}\gamma\alpha^2 = \delta \leq 1,$$

which in particular means that  $\gamma \leq N^{-\frac{1}{2}}$  the following inequality holds

$$\mathbf{P}(\sup_{x \in S_2(\Sigma_{2m})} |\frac{1}{N} \sum_{i=1}^N \langle Y_i, x \rangle^2 - 1| > \delta) \leq \exp(-(\log 2)\gamma^2).$$

## References

- [1] CANDÈS, E.J. (2008), The restricted isometry property and its implications for compressed sensing. *C.R.Math.Acad.Sci.Paris*, **346**, N.9-10, 589-592.
- [2] MENDELSON, S. PAJOR, A AND TOMCZAK-JAEGERMANN (2008) Reconstruction and sub-gaussian operators in asymptotic geometric analysis, *Geom. Funct. Anal.* **17** N. 4, 1248-1282.
- [3] MENDELSON, S, POURIS, G (2011) On generic chaining and the smallest singular value of random matrices with heavy tails. Preprint.
- [4] TALAGRAND, M. (2005), The generic chaining. *Springer-Verlag*.